ARI Contractor Report 2004-02

# Working Memory Influences on Long-Term Memory and Comprehension

**Gabriel A. Radvansky**
University of Notre Dame



**January 2004**

**United States Army Research Institute for the Behavioral and Social Sciences**

# REPORT DOCUMENTATION PAGE

| | | |
|---|---|---|
| 1. REPORT DATE (dd-mm-yy)<br>January 2004 | 2. REPORT TYPE<br>Final | 3. DATES COVERED (from. . . to)<br>June 1999 – June 2002 |
| 4. TITLE AND SUBTITLE<br>Working Memory Influences on Long-Term Memory and Comprehension | | 5a. CONTRACT OR GRANT NUMBER<br>DASW01-99-K-0001 |
| | | 5b. PROGRAM ELEMENT NUMBER<br>61102A |
| 6. AUTHOR(S)<br>Gabriel A. Radvansky | | 5c. PROJECT NUMBER<br>B74F |
| | | 5d. TASK NUMBER |
| | | 5e. WORK UNIT NUMBER<br>2901 |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>Gabriel A. Radvansky<br>University of Notre Dame<br>Notre Dame, IN 46556-5602 | | 8. PERFORMING ORGANIZATION REPORT NUMBER<br>20040130 031 |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br>U.S. Army Research Institute for the Behavioral and Social Sciences<br>5001 Eisenhower Avenue<br>ATTN: DAPE-ARI-BR<br>Alexandria, VA 22304-4841 | | 10. MONITOR ACRONYM<br>ARI |
| | | 11. MONITOR REPORT NUMBER<br>Contractor Report 2004-02 |

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

This report is published to meet legal and contractual requirements and may not meet ARI's scientific or professional standards for publication

14. ABSTRACT *(Maximum 200 words)*:

This project was conducted with the aim of understanding the role of working memory in the comprehension and long-term retention of event-specific information. This study looked at how comprehension and memory processing at the mental model level is related to traditional measures of working memory capacity, including the word span, reading span, operation span, and spatial span tests. Issues of particular interest were the ability to remember event descriptions, the detection and memory of functional relations, the detection of inconsistencies, sensitivity to causal connectivity, and memory for surface form, textbase and situation-specific content. Although traditional working memory span measures were related to a few of our tests, the relation was primarily confined to the textbase level of processing. There was little evidence that traditional measures of working memory span were directly related to processing at the mental model level.

15. SUBJECT TERMS

comprehension memory capacity memory span tests

| SECURITY CLASSIFICATION OF | | | 19. LIMITATION OF ABSTRACT | 20. NUMBER OF PAGES | 21. RESPONSIBLE PERSON (Name and Telephone Number) |
|---|---|---|---|---|---|
| 16. REPORT<br>Unclassified | 17. ABSTRACT<br>Unclassified | 18. THIS PAGE<br>Unclassified | Unlimited | 24 | Paul A. Gade<br>703-617-8866 |

Abstract

This project was conducted with the aim of understanding the role of working memory in the comprehension and long-term retention of event-specific information. This study looked at how comprehension and memory processing at the mental model level is related to traditional measures of working memory capacity, including the word span, reading span, operation span, and spatial span tests. Issues of particular interest were the ability to remember event descriptions, the detection and memory of functional relations, the detection of inconsistencies, sensitivity to causal connectivity, and memory for surface form, textbase and situation-specific content. Although traditional working memory span measures were related to a few of our tests, the relation was primarily confined to the textbase level of processing. There was little evidence that traditional measures of working memory span were directly related to processing at the mental model level.

## Working Memory Influences on Long-Term Memory and Comprehension

An important need for effective thinking and reasoning is for complex sets of information about the world to be readily understood by people. To understand information well, a person must be able to successfully comprehend the nature of the events that are being referred to by the descriptions that have been provided. The simple retention of a collection of individual pieces of information is insufficient. A person must be able to properly relate that information to external circumstances. This project sought to assess how an individual's cognitive abilities can be used to predict their future performance on comprehension and long-term memory tasks and to develop methods of training that are tailor-made to an individual's weaknesses. This project focused on working memory influences on higher-level comprehension, and the long-term memories that result from this comprehension (often called mental models because they are mental simulations of a set of circumstances that could exist in the world). Working memory is that portion of cognition where information is actively manipulated. This project assessed the degree to which the successful processing of mental models are predicted by traditional measures of cognitive ability, or can be better predicted by measures aimed more directly at how these representations are created and retrieved. This is important because our knowledge about events in the world relies more on the mental models we create than on lower level representations, such as propositional codes.

### Large-scale studies

One consistent claim from research on working memory is that people who have a larger working memory capacity are better at language comprehension and memory (e.g., Daneman & Merikle, 1996). However, many of these studies have not looked at the influence of capacity on mental model processing. Mental models (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998; Johnson-Laird, 1983) are complex mental representations that can simulate described situations. Because mental model processing requires the active manipulation of information, working memory is intimately involved. How is the processing information at the mental model level is affected by individual differences in working memory.

When people understand language, they create three types of representations. These are the surface form, the textbase, and the mental model (van Dijk & Kintsch, 1983). The surface form is a verbatim representation that is relatively short-lived (Sachs, 1967). At a deeper level is the textbase. This is an abstract representation of the idea units, or propositions. The textbase is more durable, and can be more easily retrieved from memory. Finally, further removed is the mental model. This is a representation of the situation described by a text, not the text itself (Glenberg, Meyer & Lindem, 1986). The focus here was on several aspects of mental model processing, including the processing and remembering of functional relations, the detection of inconsistencies, and memory for information about a described event. We assessed individual differences using traditional measures of working memory capacity.

#### Working memory capacity

Working memory is often operationalized in terms of capacity or "span". That is, the number of items that can be held over a period of time with the additional load of a second task,

such as reading or solving math problems. There are a number of studies relating working memory span to language comprehension and memory. The basic idea is that people with greater working memory capacity are better at a number of comprehension and memory tasks. What working memory spans actually measure is a matter of some debate (see Miyake, 2001). The traditional view is that span scores reflect how much information a person can maintain during processing. This assumes some sort of resource metaphor (Baddeley, 1986). Others have suggested that span scores reflect attentional control mechanisms, such as inhibition (Conway & Engle, 1994; Kane, Bleckley, Conway & Engle, 2001) or the management of interference (Lustig, May , & Hasher, 2001; May, Kane & Hasher, 1999). Regardless of the view taken, the focus in working memory span measures is on the retention of certain elements during active processing.

A number of working memory span tests have been developed. The most basic are <u>simple span tasks</u>, such as <u>digit span</u> or <u>word span</u>. In these tasks, people are given sets of items (i.e., digits or words, respectively) and are tested to see what is the largest set size that can be recalled. Daneman and Merikle (1996) have shown that the word span is a superior measure of language processing over the digit span. Because our concern is with language comprehension and memory, we focus on the word span. More recently, the focus has been on measures that include a processing task in addition to the need to hold a set of items in memory. Tasks with both processing and storage components are referred to as <u>complex span tasks</u>. We consider three such tasks here. One of the more popular is the Daneman and Carpenter (1980) <u>reading span</u> test. This is regarded as a measure of working memory span for language processing. In this task people read aloud a set of sentences, such as "His head sat on his shoulders like a pear on a dish." After each set, people are to report back the last word of each of the sentences in the most recent set. Performance on this measure is correlated with a number of measures of language processing and memory (see Daneman & Merikle, 1996 for a review).

Another measure is the <u>operation span</u> test, developed by Turner and Engle (1989). This task has been promoted as being more domain independent. For this measure, people are given a two-operation math problem (e.g., $(7 \times 1) + 8 = 16$) along with a word (e.g., horse). The problem and solution are read aloud and the person indicates whether the solution is correct or not. Then the word is read aloud. At the end of each set, a person recalls as many of the words from that set as they can. Performance on this measure has been related to language processing (e.g., Engle, Cantor & Carullo, 1992) and memory retrieval (e.g., Cantor & Engle, 1993).

More recently, Shah and Miyake (1996) have developed a <u>spatial span</u> test that is aimed more at the processing and retention of spatial information. The structure of this test is modeled after the reading span test. In this test people are presented with a series of rotated letters with the task of indicating whether each is normal or mirror reversed. After each set, the person reports the location of the top of each letter (among eight equally spaced locations laid out in a circle) in the most recent set. Performance on this test has been related to spatial language processing (Friedman & Miyake, 2000).

There is no question that working memory span is related to comprehension and memory. However, almost all of this research has focused on the surface form and/or textbase levels. Studies looking at memory for specific words, lexical access, and vocabulary ability are aimed at the surface form. In addition, studies looking at memory for propositional content, paraphrasing ability, and following directions are aimed at the textbase level. For studies that use standardized tests, these often emphasize the surface form, textbase, or general world knowledge. With the notable exception of a study by Friedman and Miyake (2000), which is detailed in the general

discussion, there have been very few studies that have looked at working memory capacity and its relation to processing that primarily involves the mental model level.

One of these few studies was done by Lee-Sammons and Whitney (1991) who compared memory span (as measured by the operation span test) with performance on a memory paradigm developed by Anderson and Pichert (1978). This paradigm tests how reader perspective affects text memory. Specifically, people read a text that gave a description of a house. Half of the people were asked to read from the perspective of a home buyer, and half from the perspective of a burglar. After reading, people were given two recall tests. The first simply asked people to recall the text. Importantly, on the second recall, people were asked to adopt either the original or an alternative perspective. Recalls were scored based on their propositional content. Span had no influence on memory when there was no shift in perspective. However, when there was a shift, the greater a person's span score, the more propositions that were recalled.

This study is interesting because perspective is more a quality of the mental model. One interpretation is that people with greater spans were better able to switch perspectives and recover knowledge that is more pertinent to building a new mental model. If so, this would demonstrate that working memory capacity is related to mental model processing. However, it is not clear that a perspective shift necessarily results in people creating a new situation model. It may only serve to discredit the original model. If so, people would be less likely to use the mental model during recall, and rely more on the textbase. Under these circumstances these results would reflect differences in memory for the textbase, not the mental model.

Other suggestive studies have looked at inference verification. Work by Dixon, LeFevre, and Twilley (1988) and Masson and Miller (1983) compared performance on a number of language processing and working memory tasks. The language tasks of most interest here had people verify inferences that required both information from the text along with general world knowledge (Dixon et al.) or the combination of text elements (Masson & Miller). Performance in these studies was related to reading span scores. However, it is important to note that these reading span tests departed from the standard Daneman and Carpenter (1980) test. Rather than simply reading the sentences, people also indicated whether the sentences made sense (Dixon et al.) or performed a cloze task for fragmented sentences (Masson & Miller). These tests require more complex processing that might involve mental models, whereas simply reading aloud may not. As such, it is unclear the degree to which these tasks measure working memory capacity as it has been operationalized.

## Mental model use

This section considers a number of mental model processing measures. This includes memory for general event descriptions, the comprehension and memory of functional aspects of a text, the detection of situational inconsistencies, the influence of causal connectivity on reading, and memory for situation-specific information.

Situation Identification. As stated earlier, a mental model represents the situation described by a text, not the text itself. When using mental models to make memory decisions, people are able to identify information that is consistent with a previously described situation even if that explicit information had never been encountered before. Sometimes using mental models to guide memory can result in people selecting information that is consistent with a previous description instead of selecting the actual description (e.g., Garnham, 1981; Radvansky, Gerard, Zacks, & Hasher, 1990). For our purposes, we wanted a method that required people to

use mental models to identify statements that are consistent with previous descriptions, but which have not been read before. To this end, we used a situation identification test in which people first read a series of sentences. Afterward, people were presented with a test in which they were to select one of six options that best described the same situation as the original sentence. The ability to do this was used as a measure of mental model memory use.

If working memory span is related to mental model processing, people with higher span scores should have higher situation identification scores. Having greater capacity would allow them to encode and store information in long-term memory more effectively in general. This would be reflected in an increased ability to remember what was read earlier and reason through which of the alternatives best matched the original.

Functionality. An important role of the mental model is to represent the functional relations among entities. Functional relations are those that involve the meaningful interaction between entities and which serve to define the event. These relations often convey a typical interaction between two entities, such as between a hammer and a nail. However, it may be possible for a functional relation to be atypical, such as using a rock to pound a nail. These atypical, but functional, relations are identified only when it is possible for the entities to meaningfully interact, typically based on the affordances of the entities to one another (e.g., Glenberg, 1997). Functional relations are important for metnal model processing. People find it easier to encode and remember functional as compared to nonfunctional spatial relations (Radvansky & Copeland, 2000). In addition, people are able to identify pictures of objects more quickly when the orientation of the pictured object matched how an object was described as functioning in a particular context (Stanfield & Zwaan, 2001).

If working memory capacity is related to mental model processing, a reasonable expectation is that people with higher span scores will show larger functionality effects. Their greater capacity would allow them to access and use more information from long-term memory. With more information available about the referents in the text, it would be easier to note how the elements in a situation are interacting. Thus, high span people would be in a better position to take advantage of functional relations.

Inconsistencies. If a person comprehends a text sufficiently, any inconsistencies that are present should be noticed. People are sensitive to situational inconsistencies. In a study O'Brien and Albrecht (1992) people read texts in which the location of a character was described. For example, a sentence might be "As Kim stood inside/outside the health club she felt a little sluggish" where Kim's location is varied between conditions. A later sentence in the text would be "She decided to go outside and stretch her legs a little". If Kim was initially described as being inside, this second sentence would be consistent with the previously described situation. However, if she was already outside the health club, this second sentence would be inconsistent with the described events. Reading times for critical sentences are slower in the inconsistent than the consistent condition. People can be sensitive to the internal consistency of the described situation.

Working memory span may be involved in this process with people with larger capacities showing greater inconsistency effects. The detection of inconsistencies requires a person to have available both the current information in the text, and the information from the earlier portion of the text that makes it inconsistent. The greater a person's working memory span, the more likely this information would be available.

Causal Connectivity. One of the characteristics of described situations that is not explicitly conveyed in the structure of a text itself are causal relations. There is some variability

in the degree to which various elements are causally related to one another. This is referred to as causal connectivity. The more causal connections there are, the greater the connectivity. Causal connectivity influences comprehension and memory. Specifically, information that is integrated into the causal chain of events, and is higher in causal connectivity, is better remembered (Trabasso & van den Broek, 1985), and is rated as being more important (Trabasso & Sperry, 1985). The current study uses an approach in which the elements of a text, such as clauses or sentences, are first coded in terms of the degree of causal connectivity. Then reading times for these elements are analyzed in a regression analysis with the number of causal connections as a predictor variable.

If working memory span is involved in this process, people with larger capacities will show a greater influence of causal connectivity. To be influenced by causal connectivity requires that a person have available those previous portions of the text that are causally relevant to the information that is currently being read. The greater a person's working memory span, the more likely this information is available.

Memory for Event Specific Information. As described earlier, people typically create three levels of representation -- the surface form, the textbase, and the situation model. People may vary in the degree to which they emphasize these different representations. As such, it makes sense to separate out these components using a method developed by Schmalhofer and Glavanov (1986). In this paradigm, people read a text and are then given a series of probe sentences. The task is to indicate whether a sentence was read before. There are four classes of probes: verbatims, paraphrases, inferences, and incorrects. Using signal detection analysis, the rate of responding "yes" to each of these probe types can be used to determine the strength of the three representational levels (see the method section for a more detailed description of this process).

## Study

Traditional measures of working memory span are related to comprehension and memory at the surface form and textbase levels. However, the relation to the mental model level is not clear. It may be that working memory span is a general-purpose resource. If so, one would expect it to be related to mental model processing as well. The greater one's capacity, the better one is able to construct, and therefore remember, a coherent mental model of the described events. Alternatively, it may be that traditional measures of working memory span are more confined to surface form and textbase levels of comprehension and memory, and do not have a strong, direct relation to the mental model level.

## Method

### Participants

We tested 160 participants. These people were drawn from the subject pools at the University of Notre Dame and Indiana University South Bend and were reimbursed with partial class credit. All were native English speakers. The data from one person was excluded for having a situation identification score at chance (4 out of 24) as well as very short reading times, suggesting that he was not actually reading the stories.

### Materials and procedure

Memory span tests. There were four tests of working memory capacity. The most basic was a word span test. For this test, people were presented with a series of words in ever increasing set sizes of 3 to 8, with 3 sets at each set size, and an additional 3 practice trials at set size 3. The words were presented one at a time on a computer screen for 1 s each. At the end of

each set, a series of question marks was displayed, one for each item in the set. The task was to recall the words in the order that they were presented. People responded by typing their responses into the computer. People typed "DK" ("Don't Know") for those items they could not recall. The order of the words was randomized for each participant.

For the Daneman and Carpenter (1980) reading span test people were presented with a series of sentences in ever increasing sets of 2 to 6, with 5 sets at each set size, except for set size six, for which there were only three sets. The sentences were presented on a computer screen and the task was to read each sentence aloud. After finishing each sentence, the experimenter advanced to the next sentence by pressing the space bar. When a blank screen appeared, the task was to recall the final word from each of the sentences in the current set. The experimenter typed these responses into the computer. Again, people were encouraged to recall these in the order that they were presented. The order of the sentences was randomized for each participant.

For the Turner and Engle (1989) operation span test people were presented with a series of math problems followed by a word, such as "(9 X 1) + 1 = 9" then "BOAT" in ever increasing sets of 2 to 7, with 3 sets at each set size, except for set size 2, for which there were 6 sets (the first three were considered practice). The problems and words were presented on a computer screen, and the task was to read each problem aloud and indicate whether the answer was correct by pressing one of two buttons on a computer mouse, and then read the word aloud. When a blank screen appeared, the task was to recall the words from that set. The experimenter typed these responses into the computer. Again, people were encouraged to recall these in the order in which they were presented. The order of the problems and words was randomized for each person.

For the Shah and Miyake (1996) spatial span test people were presented with a series of rotated letters (i.e., F, J, L, P, & R) in ever increasing sets of 2 to 6, with 5 sets at each set size, except for only 3 at set size 6. The letters were displayed on index cards, and the processing task was to state whether the letter was normal or mirror-imaged. The responses were recorded by the experimenter. After each set, a card was presented with eight locations mapped out around a circle at 45 degree intervals (0 degrees, or upright, was never used). The task was to indicate the position of the top of the letters in the current set in the order in which they were encountered. Note that the same location was never repeated within a set. The experimenter recorded these responses. People were encouraged to guess if they were unsure.

For the scoring of all of the span tests, if a set was recalled correctly, then the score was equal to the number of words in that set. The scores for all of the correctly recalled sets were then totaled (Conway & Engle, 1994).

Situation identification test. In this task, people were given a series of 24 sentences, which were drawn from or inspired by those materials used by Garnham (1981). These sentences were presented one at a time on a computer screen. During this initial presentation, the task was to rate the sentences for pleasantness. Responses were entered into the computer using a 1 to 7 scale, with 1 indicating "extremely unpleasant" and 7 indicating "extremely pleasant". Because this was a cover task, these ratings were not recorded.

After all the sentences were rated, people were given a surprise identification task. On each trial, six alternatives were presented which were variations of the original sentence. These six alternatives were presented, in a random order on each trial for each person. The task was to select the item that most closely described the same situation as the sentence read and rated earlier. People were informed that the original sentences would never appear and to type their responses into the computer. No feedback was provided. The six alternatives were (a) altered

prepositional phrase, (b) altered direct object, (c) altered verb, (d) altered prepositional phrase and direct object, (e) altered prepositional phrase and verb, and (f) altered direct object and verb. An example of six sentences for the original sentence "The man lost a hand of poker at the card shark's." are presented below. Sentence 2 is the correct response.

1. The man lost a hand of poker like the card shark.
2. The man lost some money at the card shark's.
3. The man won a hand of poker at the card shark's.
4. The man lost some money like the card shark.
5. The man won a hand of poker like the card shark.
6. The man won some money at the card shark's.

Reading and memory task. For this study people were asked to read a series of eight narratives. These stories were 31 to 45 sentences in length ($\underline{M} = 38$). An example of one of these texts is presented in the Appendix. These stories were presented one clause at a time on a PC-compatible computer in white on a black background. People pressed a space bar with their left hand to advance to the next clause. Reading times were collected. People rested their right hand on the computer mouse to respond to the comprehension questions at the end of each story. The left button was marked with a "Y" for "yes" and the right button was marked with an "N" for "no". A number of measures were gathered during story reading. There were also a few memory measures that were presented using paper and pencil after all of the stories were read. We will discuss each of these in turn.

Functionality. To assess sensitivity to functional relations, each story contained two sentences that described a spatial relation. For two of the stories, both of the relations were functional. For two stories, both were nonfunctional. Finally, for four stories one was functional and the other nonfunctional. The combinations of conditions to stories were rotated across participants. To illustrate the difference between the functional and nonfunctional versions, in the story in the Appendix, the functional version was "David was standing below an old bridge." This is functional because it allows David to get out of the rain. The nonfunctional version was "David was standing next to a lamppost." This is nonfunctional because this spatial relation plays no role in the story. In all cases, the functional version was changed to a nonfunctional version by altering both the spatial relation and the reference object. To assess whether functionality had an effect on comprehension, we used the reading time for these sentences. Previous research has shown that functional sentences are read more quickly than nonfunctional sentences (Radvansky & Copeland, 2000). Presumably, people are influenced by how well the information fits into the causal chain of the described events.

In addition, after people had read all of the stories they were given a recognition test for this functional information. People were presented with each original sentence with three distractors. The distractors altered the located object in the sentence, the spatial relation between the two or both. For example, if the original sentence was "David was standing below an old bridge", the distractor sentences would be "David was standing next to an old bridge." "David was standing below a lamppost." "David was standing next to a lamppost." Thus, the functional and nonfunctional versions were always present. These options were randomized and the order in which the various trials occurred was also randomized. People responded by circling the letter next to each sentence (a, b, c, or d) that they thought corresponded to the sentence they had read earlier.

Inconsistency. To assess inconsistency detection, in the texts were three sentences that described actions that were either consistent or inconsistent with the current location. Each story

had either 1 consistent and 2 inconsistent, or 2 consistent and 1 inconsistent sentences. The versions were rotated in each story across participants. To illustrate the difference between consistent and inconsistent sentence versions, in the story presented in the Appendix there is a statement that David is walking along the banks of a river. Several sentences later, a sentence occurs that is either consistent ("David walked further down the river.") or inconsistent ("David walked outside to the river.") with the previously described location. Reading time for this target sentence was recorded and used as the dependent measure. Across all stories there were 24 consistency sentences.

Causal Connectivity. To assess causal connectivity, we coded the experimental stories' causal structure. We scored two story clauses as being causally connected if they met the following criteria: (a) the first event had started prior to the second event, (b) the first event was still in operation when the second event started, (c) the occurrence of the first event was necessary for the second event, and (d) the first event was sufficient for the second event, either by itself or in conjunction with other narrative events. As an example from the story in the Appendix, the sentence "The nights could be bitterly cold" causes "He hugged his thin dust-covered coat around him". In addition to this standard scoring procedure, we also scored two narrative events as being causally related, but in a backwards manner, if the second event preceded the first.

After the narratives were scored, we used causal connectivity as a predictor variable in a regression analysis with reading times as the dependent variable. The idea was that if causal connectivity is important, then the more causal connections there are, the easier it should be to read, and the faster the reading time. In addition to causal connectivity, we included a number of text variables in the regression analyses that are known to affect reading times. These were (a) number of syllables, (b) word frequency, (c) serial position, and (d) new arguments. Reading times for the inconsistency and functionality items were excluded form this analysis because their relation to the causal structure of the text varied depending on which version a person read.

Comprehension Questions. Although not aimed at a particular level of processing, after each story, a pair of general comprehension questions were presented, such as "Did the travel agent lie to David?" and "Did David like the coffee at the hotel?" Their purpose was to encourage people to actively read the texts. These comprehension questions required yes or no responses that were recorded by having readers press one of two buttons on the computer mouse.

Levels of representation. Using the Schmalhofer and Glavanov (1986) paradigm, after reading people were given a recognition test. Sixteen sentences were selected from each text. There were four types of probes. Verbatim probes were sentences that had appeared in the text (e.g., "The driver filled up the gas tank"). Paraphrase probes contain the same propositions, but expressed differently. Sentences were reworded so that they retained the propositional content of the original. This was done using synonyms or altering word order (e.g., "The gas tank was being filled by the driver"). Inference probes were information that was not mentioned, but which was consistent with and important to the described situation (e.g., "The gas tank of the car was near empty"). Finally, incorrect probes were composed of information that was neither mentioned and was unlikely to be inferred. However, the information was globally consistent with the passage's theme (e.g., "The driver washed the windows of the car").

The task was to indicate whether a sentence had been read earlier. People were warned that the sentences might contain slight wording changes. The probes were blocked by story, with the title appearing prior to the probes. This was done so that people knew to which passage the items referred. The story order was the same as during reading. The order of the probes within a

story block was randomized. People responded by circling either a "Y" or an "N" located next to the probe sentence to indicate whether it was an old or new sentence. The type of probe (i.e., verbatim, paraphrase, inference, or wrong) for a particular sentence was rotated across participants.

The ability to discriminate verbatim from paraphrase probes is an index of the surface representation. The difference between these is that one matches the original surface characteristics and the other does not. They are equivalent in how they map onto the textbase and mental model. Similarly, the ability to discriminate between paraphrase and inference probes is an index of the textbase. The difference between these is that one matches the original in propositional content and the other does not. Also, both are inconsistent with the surface structure and consistent with the described situation. Finally, the ability to discriminate between inference and incorrect probes provides an index of the use of mental models. The difference between these is that one corresponds to the described situation and the other does not. Both are inconsistent with the surface structure and textbase.

*A'* scores (following Donaldson, 1992), a signal detection measure, were calculated as discrimination measures. For the surface form measure, verbatims were considered hits and paraphrases were considered false alarms. For the textbase measure, paraphrases were hits and inferences were false alarms. Finally, for the mental model measure, inferences were hits and incorrects were false alarms.

Data treatment. All of the reading time data, including the functionality and inconsistency items, were trimmed by first eliminating any clearly deviate times (less than 50 ms / syllable or greater than 1500 ms / syllable). In addition, for sentences in the functional and inconsistency analyses, the fastest and slowest reading times per condition per participant were dropped (cf. Rinck & Bower, 1995).

## Results

As an overview, for all of our measures of mental model processing, the expected effects were observed. However, there was no clear evidence that performance on the working memory span tests was related to comprehension and memory at the mental model level. Instead, memory span seemed to be more related to the textbase level.

Working Memory Span. Summary data for these tests are presented in Table 1. Moreover, the correlations between the various span measures are presented in Table 2. As can be seen, the span tests were moderately correlated with one another.

Situation Identification. The situation identification test scores ranged from 7 to 23 (out of 24 possible), with a mean of 17.7 ($SD$ = 3.3). Thus, people could make judgments about the described situation fairly well based on their memory of what they had read earlier. The correlations with the various span scores are presented in Table 3. As can be seen, there was no significant relationship with any of these measures.

Functionality. None of the reading time data were trimmed for exceeding the long criterion, but 1.7% of the data were trimmed for being too fast (< 50 ms/ syllable). Following this, the fastest and slowest reading times in each condition were trimmed as described in the method section. Overall, people read the spatial relation sentences faster when they conveyed a functional relation (181 ms/syllable) than a nonfunctional relation (198 ms/syllable), $F(1,158) = 33.86$, $MSe = 658$, $p < .001$. This is consistent with previous research (Radvansky & Copeland, 2000). The difference between the reading times in the two conditions was used as an index of the functionality effect to compare with the memory span tests. The results of the correlation

analyses are presented in Table 3. As can be seen, there was no significant relation with any of these measures. If anything, most of the correlations are nominally in the wrong direction.

For the recognition data, people identified the sentences better when they conveyed a functional relation (82%) than a nonfunctional relation (57%), $\underline{F}(1,158) = 214.17$, $\underline{MSe} = .024$, $\underline{p} < .001$. The difference between the recognition rates in the two conditions was used as an index of the functionality effect to compare with the memory span tests. The results of the correlation analyses are presented in Table 3. As can be seen, there was significant relationship with the reading span test, but none of the other measures. Moreover, the direction of this relationship is the opposite of what was predicted. The greater the reading span score, the smaller the functionality effect.

Closer inspection of the recognition data revealed an interesting relationship between the memory span measures and performance on the functional and nonfunctional items. The pertinent analyses are presented in Table 4. Specifically, the memory span scores were related to performance on the nonfunctional items. In contrast, there was no relationship with the functional items. This is consistent with the idea that span tests tap more into propositonal memories, not the mental model level.

Inconsistencies. Overall, people read consistent sentences faster (195 ms/syllable) than inconsistent ones (220 ms/syllable), $\underline{F}(1,158) = 67.09$, $\underline{MSe} = 747$, $\underline{p} < .001$. Thus, comprehension was disrupted by information that was inconsistent with the current state of affairs. This is consistent with previous research (e.g., O'Brien & Albrecht, 1992). The difference between these two reading times was used as an index of inconsistency to compare with the memory span tests. The results of the correlation analyses are presented in Table 3. As can be seen, there was no significant relation with any of these measures. Again, if anything, the correlations were nominally in the wrong direction.

Causal Connectivity. The degree of causal connectivity was significantly related to reading time, with a mean beta-weight of -.049, $\underline{t}(159) = -15.79$. Thus, people were sensitive to the causal structure of the texts, as predicted by mental model theory. The causal beta-weight from the regression analyses was used as an index of sensitivity to causal connectivity to compare with the memory span tests. The results of the correlation analyses are presented in Table 3. As can be seen, there was no significant relation with any of these measures, with the correlations nominally in the wrong direction.

Comprehension Questions. Overall, performance on the comprehension questions was good, at 93% correct. As can be seen in Table 3, only the operation span test was significantly related to performance. It should be noted that these questions asked about detail information that was not necessarily information that would have been retained in the situation model alone. The surface form and textbase levels of representation could have been used as well. Therefore, some influence of propositional memory would be expected although there is also a mental model component.

Schmalhofer and Glavanov analysis. The recognition test data revealed higher A' discrimination scores at the situation model level (mean A' = .79), less so at the textbase level (A' =.68), and lowest at the surface form level (A' = .58). All of these values are significantly different from one another, all $\underline{p}s < .001$, and all were significantly greater than chance, all $\underline{p}s < .001$. Thus, in this case, memory for the texts was dominated by the mental model level.

As can be seen in Table 3, not surprisingly because they were so close to chance (.5), surface form A's were unrelated to all of the span measures. However, for the textbase A's, performance were significantly related to the operation span task, and marginally so to the spatial

span task. The correlations with the other two span measures did not reach significance. Finally, the mental model A's were not significantly related to any of the span measures, and if anything, they were consistently in the wrong direction.

## Discussion

The current study looked at performance on a variety of mental model comprehension and memory processes and their relation to standard working memory span measures. Memory span was found to be related to memory for functional relations, general comprehension question accuracy, and textbase memory. However, even in these cases, the relation was limited to specific span tests, with no strong pattern emerging. On the whole, there was very little observed relation between working memory span and performance on mental model level tasks.

Memory span tests may be more sensitive to processing at the textbase level than at the situation model level. Performance on the textbase A' measure and general comprehension questions clearly involves this level of representation. Also, memory for functional relations could also be viewed as involving a textbase component because this task involves recognition of what was actually in the text. This idea is further reinforced by the finding that span is related only to the nonfunctional items, which are generally remembered more poorly overall. This nonfunctional information would not be as well integrated into the mental model, and thus, performance on these items would be more a reflection of textbase memory.

Let's consider the memory span tests. For the <u>word span</u> test, although it was rarely better than the others, its relation to performance was generally consistent with the more complex span measures. The <u>reading span</u> test has been argued to be a measure of general language processing ability (Daneman & Carpenter, 1980). As such, it would be expected that this measure might fare better than most of the others. However, this was not what was observed. It was only significantly correlated to a single performance measure, the recognition of spatial descriptions, and then only to the nonfunctional ones, like all of the other span tests. The <u>operation span</u> test has been promoted as a measure of more general-purpose cognition (Kane, Bleckley, Conway, & Engle, 2001). As such, it might be a better measure of general cognitive ability. Consistent with this idea, this measure was related to two performance measures, comprehension question accuracy and textbase A', whereas the other span tests were related to only one measure, at best. If working memory span tests are best viewed as measures of textbase processing, the operation span score does this task better than the others.

Finally, the <u>spatial span</u> test has been put forward as tapping into spatial processing (Shah & Miyake, 1996). Moreover, Friedman and Miyake (2000) have suggested that this measure also picks up on the processing of spatial information during language comprehension. This is important in the context of the current study because many of our mental model measures involved spatial information. The functionality measures were assessing spatial functionality. The inconsistency measure tapped into spatial inconsistencies. However, in all of these cases, spatial span did no better than the other span tests. The only time spatial span distinguished itself from the others was on the situation model A' measure. However, even here, the relation is in the opposite direction of what would be expected if spatial working memory were important. The pattern of results is consistent with the idea that even this span test reflects lower level processing.

Although, it appears that traditional memory span measures are more in tune with processing at the textbase level than the situation model level, this is not to say that span is unimportant for language comprehension. Obviously it is. All we are saying is that these indices

appear to be measuring processing at lower or intermediate levels. Furthermore, we are not trying to argue that memory span has no implications for mental model processing. If processing is sufficiently disrupted at lower levels, this will complicate the ability to create coherent and accurate situation models that are built from this information. Thus, memory span may have indirect influences on mental model processing that were not observed here.

Mental Model Dimensions and Working Memory

In a pair of experiments by Friedman and Miyake (2000), performance on reading span and spatial span tests were compared with mental model processing during reading. That study focused on causality and space. People read texts that described characters moving about in buildings. During reading, they were interrupted with probes to assess either causal or spatial information. The causal probes were inferences. The spatial probes were maps of the building with one of the rooms highlighted. The task was to indicate whether that room was the protagonist's current location.

Friedman and Miyake (2000) interpreted their results as showing that the spatial and causal dimensions were processed using different components of working memory, with the reading span being more related to the causal dimension and the spatial span being more related to the spatial dimension. Performance seemed to vary as a function of the amount of capacity in the verbal and spatial portions of working memory for each of these dimensions, respectively.

However, a closer examination of the data calls this interpretation into question. For the causal probes, reading span was correlated with accuracy, but not response times in the first experiment, and with neither in the second experiment. Moreover, although causal processing was unrelated to spatial span in the first experiment, it was related to spatial accuracy in the second experiment unless two outliers were removed. Thus, the relation between reading span and causal processing was weak. Furthermore, it is unclear to what degree this causal task depended on the retrieval of textbase information. It is possible that the significant relation between reading span and causal probe accuracy in the one experiment is tapping into the textbase, not the mental model level.

For the spatial probes, in both experiments spatial span was not related to accuracy but was related to response time. Moreover, in the first experiment, spatial probe accuracy was related to reading span, unless one outlier was removed. In the second experiment, spatial probe response time was related to reading span, unless two outliers were removed. The evidence for the relation between spatial span and situation model processing is better here, but it is confined to the response time data. However, it is unclear the extent to which this task taps mental model processing. Probe presentation disrupts comprehension and requires people to do an explicitly spatial task involving a map. It may be that this task leads people to coordinate information in the situation model with a more context independent mental map. Thus the spatial span-response time relation may reflect mental map, not situation model, processing. Given the absence of a relation between spatial span and the mental model measures in the current study, it is plausible that Friedman and Miyake's (2000) results were due more to the nature of the probe task than to mental model processing per se.

## Conclusion

The current study showed that working memory capacity, as it has traditionally been operationalized, is a good predictor of success of some of the cognitive processes involved in language comprehension and memory. However, it does not have the broad scope that some researchers have ascribed to it. Indeed it seems to be confined to lower levels of processing, such

as the textbase. There are many aspects of comprehension and memory that are unrelated to memory span, including many processes involved in cognition at the mental model level.

## References


Anderson, R.C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. Journal of Verbal Learning and Verbal Behavior, 17, 1-12.

Baddeley, A. D. (1986). Working Memory. Oxford, England: Oxford University Press.

Cantor, J. & Engle, R. W. (1993). Working-memory capacity as long-term memory activation: An individual differences approach. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 1101-1114.

Conway, R. A. & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. Journal of Experimental Psychology: General, 123, 354-373.

Copeland, D. E., & Radvansky, G. A. (2002). Working memory and syllogistic reasoning. Manuscript submitted for publication.

Daneman, M, & Carpenter, P. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

Daneman, M, & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. Journal of Experimental Psychology: General, 130, 208-223.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. Psychonomic Bulletin & Review, 3, 422-433.

Dixon, P., LeFevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. Journal of Educational Psychology, 80, 465-472.

Donaldson, W. (1992). Measuring recognition memory. Journal of Experimental Psychology: General, 121, 275-277.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. Journal of Experimental Psychology: Learning, Memory and Cognition, 18, 972-992.

Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: Recognition memory for three types of textual information. Discourse Processes, 13, 175-190.

Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. Journal of Experimental Psychology: General, 129, 61-83.

Garnham, A. (1981). Mental models as representations of text. Memory & Cognition, 9, 560-565.

Glenberg, A. M. (1997). What memory is for. Behavioral and Brain Sciences, 20, 1-55.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. Journal of Memory and Language, 26, 69-83.

Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. Journal of Experimental Psychology: General, 108, 151-181.

Johnson-Laird, P. N. (1983). Mental Models: Towards a cognitive science of language, inference and consciousness. Cambridge, MA.: Harvard University Press.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory. Journal of Experimental Psychology: General, 130, 169-183.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. Journal of Memory and Language, 30, 580-602.

Kintsch, W., Welsch, D., Schmalhofer, F. & Zimny, S. (1990). Sentence memory: A theoretical analysis. Journal of Memory and Language, 29, 133-159.

Lee-Sammons, W. H., & Whitney, P. (1991). Reading perspectives and memory for text: An individual differences analysis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 1074-1081.

Light, L. L., & Anderson, P. A. (1985). Working-memory capacity, age, and memory for discourse. Journals of Gerontology, 40, 737-747.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. Journal of Experimental Psychology: General, 130, 199-207.

Masson, M. E., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. Journal of Educational Psychology, 75, 314-318.

May, C. P., Kane, M. J., & Hasher, L. (1999). The role of interference in memory span. Memory & Cognition, 27, 759-767.

Miyake, A. (2001). Individual differences in working memory: Introduction to the special section. Journal of Experimental Psychology: General, 130, 163-168.

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. Journal of Experimental Psychology: Learning, Memory and Cognition, 18, 777-784.

Radvansky, G. A. (1999) Aging, memory and comprehension. Current Directions in Psychological Science, 8, 49-53.

Radvansky, G. A., & Copeland, D. E. (2000). Functionality and spatial relations in memory and language. Memory & Cognition, 28, 987-992.

Radvansky, G. A., & Copeland, D. E. (2001). Working memory and situation model updating. Memory & Cognition, 29, 1073-1080.

Radvansky, G. A., Copeland. D. E., & Zwaan, R. A. (in press). Aging and functional spatial relations in comprehension and memory. Psychology and Aging.

Radvansky, G. A. Gerard, L. D., Zacks, R. T., & Hasher, L. (1990). Younger and older adults' use of mental models as representations of text materials. Psychology and Aging, 5, 209-214.

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation Models and Aging. Psychology and Aging, 16, 145-160.

Rinck, M., & Bower, G. H. (1995). Anaphor resolution and the focus of attention in situation models. Journal of Memory and Language, 34, 110-131.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. Perception & Psychophysics, 2, 437-442.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. Journal of Memory and Language, 25, 279-294.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. Journal of Experimental Psychology: General, 125, 4-27.

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging inference processes. Memory & Cognition, 20, 539-548.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. Psychological Science, 12, 153-156.

Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. Journal of Memory and Language, 24, 595-611.

Trabasso, T., & van den Broek, P. W. (1985). Causal thinking and the representation of narrative events. Journal of Memory and Language, 24, 612-630.

Turner, M. L., & Engle, R. W. (1989). Is working memory task dependent? Journal of Memory and Language, 28, 127-154.

van Dijk, T. A., & Kintsch, W. (1983). Strategies in Discourse Comprehension. New York: Academic Press.

Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 920-933.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. Psychological Bulletin, 123, 162-185.

## Appendix

David walked along the banks of the river in town. Although the days were unbearably hot, the nights could be bitterly cold. He hugged his thin dust-covered coat around him as he thought about how the travel agent had lied to him. Two months ago, David went to see the travel agent. She told him that North African Sahara towns were friendly and romantic. Now, everything so far suggested the opposite.

<u>David walked further down the river</u>. (consistent)

<u>David walked outside to the river</u>. (inconsistent)

A steady cold rain began to pour from the sky. If he stayed out much longer he would get soaked.

<u>David was standing below an old bridge</u>. (functional)

<u>David was standing next to a lamppost</u>. (nonfunctional)

He listened to the rain falling on the road as he took stock of their misfortunes so far. The townsfolk treated you with contempt if you didn't speak the native Arabic or French. His wallet and passport had been stolen. Maureen and he were shocked to find out how decrepit and dirty their hotel was. Even the coffee they were served was bad. David was sure that this trip would bring his troubled marriage to an end. Twenty minutes later, David saw a taxi and hailed it.

<u>The driver stopped and David got out</u>. (consistent)

<u>The driver stopped and David got in</u>. (inconsistent)

As he was scanning the drab city he saw an object that could free him. While driving through the merchant district, he saw an old black Ford. Although it was far from perfect, He thought that he could use it to escape this cursed place. David couldn't take his eyes off that car. The driver had just pulled into a gas station.

<u>The old car was sitting to the left of a slick new gas pump</u>. (functional)

<u>The old car was sitting in front of a slick new Mercedes</u>. (nonfunctional)

The contrast was striking. The driver filled up the gas tank.

<u>David wished he had his wallet...</u> (consistent)

<u>David wished pulled out his wallet...</u> (inconsistent)

...so that he could offer to buy that car. Maybe he would just steal it. How liberating it would be to cruise out of this town in that car. He didn't know where he would drive to, he just wanted out of here. Even sitting in a gas station it seemed to command his attention.

Table 1. Summary of mental ability test scores.

| | Mean | SD | Minimum | Maximum | scoring |
|---|---|---|---|---|---|
| Word Span | 35.2 | 13.1 | 9 | 77 | count score |
| Reading Span | 22.5 | 12.6 | 4 | 69 | count score |
| Operation Span | 21.6 | 11.1 | 0 | 54 | count score |
| Spatial Span | 29.0 | 19.2 | 0 | 84 | count score |

Table 2. Correlations among the various span tests.

| | Word | Reading | Operation | Spatial |
|---|---|---|---|---|
| Word Span | 1.00 | | | |
| Reading Span | 0.59 | 1.00 | | |
| Operation Span | 0.49 | 0.61 | 1.00 | |
| Spatial Span | 0.40 | 0.55 | 0.39 | 1.00 |

All of these correlations were significant, $p < .001$, with a Bonferroni correction.

Table 3. Correlation analyses results.

| | Memory Span Tests | | | |
|---|---|---|---|---|
| | Word | Sentence | Operation | Spatial |
| Situation Identification | .05 | .11 | .07 | .14 |
| Functionality (reading times) | .04 | -.03 | -.12 | -.11 |
| Functionality (recognition accuracy) | -.15 | -.20 * | -.16 | -.16 |
| Inconsistencies (reading times) | -.08 | -.05 | -.08 | -.06 |
| Causal connectivity (beta-weight) | .00 | -.06 | -.02 | -.14 |
| Comprehension questions (accuracy) | .15 | .14 | .23 * | .08 |
| Surface form (A') | -.02 | .02 | -.09 | .08 |
| Textbase (A') | .16 | .16 | .23 * | .19 ** |
| Mental Model (A') | -.09 | -.03 | -.04 | -.16 |

* corresponds to $p < .05$ (Bonferroni corrected for each dependent variable).
** corresponds to $.05 \leq p \leq .10$

Table 4. Functionality recognition correlation analyses broken down by condition.

| | Memory Span Tests | | | |
|---|---|---|---|---|
| | Word | Sentence | Operation | Spatial |
| *Nonfunctional items* | .26 * | .28 * | .21 * | .24 * |
| *Functional items* | .14 | .11 | .05 | .10 |

* corresponds to p < .05 (Bonferroni correct for each dependent variable).